MIND, BRAIN, AND EDUCATION

Exploring the Implementation of Early Math Assessments in Kindergarten Classrooms: A Research-Practice Collaboration

Jennifer A. McDonald[1], Rebecca Merkley[2], Jacqueline Mickle[1], Lisa-Marie Collimore[1], Zachary Hawes[3], and Daniel Ansari[4]

ABSTRACT— Research in cognitive development has highlighted that early numeracy skills are associated with later math achievement, suggesting that these skills should be targeted in early math education. Here we tested whether tools used by researchers to assess mathematical thinking could be useful in the classroom. This paper describes a collaborative project between cognitive scientists and school board researchers/educators implementing numeracy screeners with kindergarten students over the course of three school years. The Give-N task (Wynn, 1990) was used with first-year kindergarten students and the Numeracy Screener (Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013) with second-year kindergarten students. Results indicated that educators (N = 59) found the tools feasible to implement and helpful for exploring their students' thinking and targeting instruction. The educators' feedback also helped inform improvements to the implementation of the tools and future directions for both the schools and the researchers. This work emphasizes the importance of transdisciplinary collaboration to address the research-practice gap.

[1]Halton Catholic District School Board
[2]Department of Cognitive Science, Carleton University
[3]Applied Psychology & Human Development, Ontario Institute for Studies in Education, University of Toronto
[4]Department of Psychology, Brain and Mind Institute, University of Western Ontario

Address correspondence to Rebecca Merkley, Department of Cognitive Science, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada, K1S 5B6. e-mail: rebecca.merkley@gmail.com

Transdisciplinary collaborations are needed to bridge the gaps between cognitive science research, education research, and pedagogical practice (Ansari & Coch, 2006; Bruce et al., 2017), yet applying findings from research to the classroom requires effort and engagement on multiple levels. Collaborative projects between researchers and educators are one promising way to address the research-practice gap in cognitive science and education (e.g., Agarwal, Bain, & Chamberlain, 2012; Amiel & Tan, 2019; Campbell & Parr, 2015; Massonnié, Frasseto, Mareschal, & Kirkham, 2020). Here, we engaged in a collaborative project between a school board and a research laboratory to address mutual interests in early math education and cognitive development. Specifically, the project investigated whether early math assessments commonly used in research could be implemented and utilized by educators in kindergarten classrooms.

Educators' Need for Evidence-based Early Math Education Resources

Identifying where children are along a developmental continuum of mathematical understanding and providing intentional, focused mathematical learning opportunities is critical to success in both mathematical and overall academic achievements (Duncan & Magnuson, 2011; Engel, Claessens, & Finch, 2013; Ginsburg, 2009). While teachers are exposed to the foundations for literacy in pre-service education, many early childhood educators and primary school teachers do not receive training in mathematics pedagogy and thus have no specific math-related qualifications

(Ginsburg, Lee, & Boyd, 2008; Youmans, Coombs, & Colgan, 2018).

Prior to this project, Curriculum and Special Education staff in a school board recognized the importance of identifying students with weak foundational math skills, but they did not have a way to determine or monitor children's numerical identification ability from an early age. They wanted to find a means to identify Kindergarten students' numerical and spatial abilities early and support Kindergarten educators in providing playful mathematical learning opportunities. School board staff therefore approached a nearby research lab focused on mathematical cognition and learning to ask whether their research tasks could be implemented in their classrooms. Research on the development of mathematical thinking can inform the development of formative assessment tools for classroom practice (Ginsburg, 2009). Formative assessment is "assessment designed to guide the teaching of students, particularly those who experience difficulty in learning" (Ginsburg, 2009, p. 109). If teachers are given opportunities to observe their students' mathematical thinking, this can help them direct their instruction.

Researchers' Need for Input from Educators

Research in cognitive science and education has identified knowledge of number symbols (i.e., count words and Arabic digits) as key foundational skills of mathematics (Merkley & Ansari, 2016; Purpura, Baroody, & Lonigan, 2013). The Give-A-Number (Give-N) task (Wynn, 1990) is widely used in developmental cognitive research to assess children's understanding of the cardinal principle, or that the last number word they say when they count represents how many items are in the set. Most children start school with a good understanding of number words, but this is influenced by the experiences they have had at home and in preschool (e.g., Maloney, Converse, Gibbs, Levine, & Beilock, 2015; Sarnecka, Negen, & Goldman, 2018). Identifying children who start school without this number knowledge could help teachers adapt their provision of early math lessons to offer more activities targeting the development of this foundational understanding in these students. However, while the Give-N task has been used extensively for research purposes, its potential utility as an applied tool has not been investigated. Similarly, the Numeracy Screener was developed by members of the participating research laboratory (Nosworthy et al., 2013). It has undergone a few revisions and been used in different school districts in Ontario. Kindergarten students' scores on this assessment tool have been shown to predict first-grade math grades (Hawes, Nosworthy, Archibald, & Ansari, 2019). While participating school board staff have given informal feedback on the implementation of the assessment tools, the lab had never systematically asked educators to evaluate whether using the tools were helpful to them and informed their practice. Therefore, we collaboratively embarked on the current project to collect feedback from educators on whether the assessment tools are useful and how the tools could be modified for classroom practice.

Current Study

Given the identified need to determine and monitor children's mathematical abilities from an early age, the school board staff decided to implement the laboratory assessment tools. These tools were selected to complement the educators' existing assessment and intervention strategies to enhance early numeracy development. Board staff partnered with the Numerical Cognition laboratory, who provided access to the assessment tools and scoring resources and supported professional development with Kindergarten teachers and Designated Early Childhood Educators (ECEs).[1] Because there was limited understanding of practical implementation of the assessment tools by the research lab and school board, we set out to conduct a collaborative inquiry into the implementation and use of the tools (Donohoo, 2013).

The school board's project team outlined a theory of action to help meet students' learning needs and to contribute to the long-term outcome of helping students reach higher levels of mathematical and overall achievement. Specifically, the team planned to obtain feedback from teachers and ECEs to explore the following inquiry questions:

1. What were educators' perceptions of implementing the Give-N task and Numeracy Screener?
2. Were the assessment tools effective in identifying specific mathematical needs of students at risk for falling behind?
3. How did educators use the assessment results to inform and target their math instruction?

This inquiry project was conducted over three consecutive school years (September–June) and the implementation of the assessment tools was revised in the second and third years based on the results from the previous iterations. Here, we report on the methods and results separately for the first two cohorts of participants. We have included the methods and results for the third cohort in Supporting Information due to the word limit of the main text. Figure 1 shows the timeline of the project.

COHORT 1

Method

Participants

The first year of the project (i.e., Cohort 1) involved eight pilot schools, with 25 kindergarten classes and 54 educators

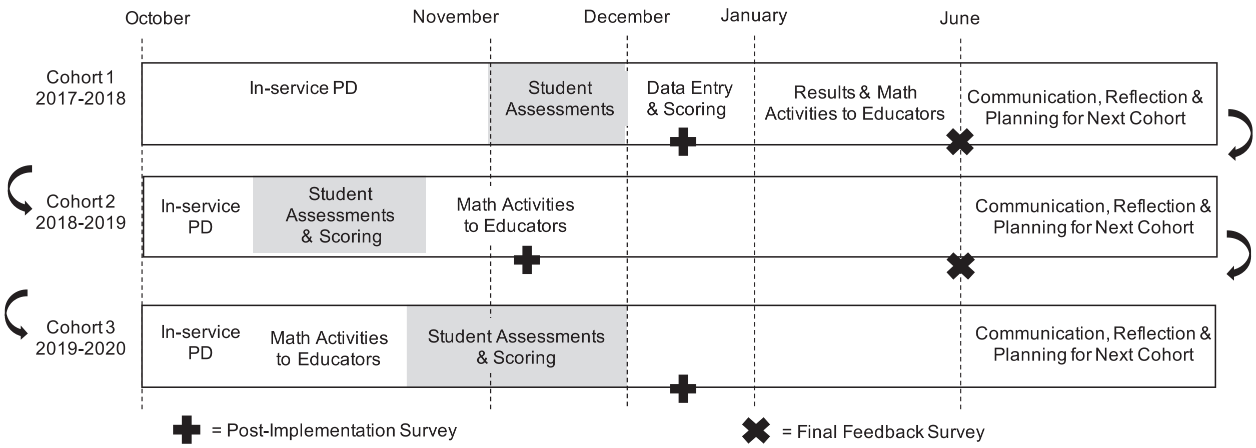

Fig. 1. Timeline of the project.

who received in-service training to administer the assessment tools. Implementation was divided among teachers and ECEs in each classroom such that ECEs typically did Give-N with the Junior Kindergarten (JK) students, and the teachers did the Numeracy Screener with the Senior Kindergarten (SK) students. Participating schools were selected based on need, as demonstrated by scores on standardized assessments, as well as regional representation and buy-in from school administrators. In total, 325 JK students and 340 SK students completed the relevant assessments ($N = 720$). In Ontario, kindergarten classes have JK students who are in the first year of the 2-year kindergarten program, and are around 4 years old, and SK students who are in the second year and around 5 years old. Demographic data were not recorded to protect the students' privacy.

Self-report data were collected from up to 31 educators (57% response rate). Participants were only asked to answer questions about the assessment tasks they were involved in administering via branching questions and may not have completed the final survey, so the number of respondents varies.

Assessment Tools

Give-N Task. JK students completed the Give-N task (Wynn, 1990). In this task, the assessor asks the child to feed a puppet sets of different numbers of objects. Children are asked to give up to eight objects, and the number asked for is adjusted based on their responses, which is known as the titration method. Children are considered an N-knower if they correctly give that number of blocks twice, and incorrectly give the next highest number twice (e.g., a child who correctly gave two blocks twice when asked for two but failed to give three blocks correctly twice when asked for three, would be considered a 2-knower). Children who can correctly give at least five blocks are considered cardinal principle-knowers (CP-knowers). It is thought that once children have learned 5, they can generalize to larger numbers and understand the cardinal principle. We adapted a scoring sheet to make the task simple for teachers to administer and score (see Supporting Information), and we provided teachers with puppets and cubes to administer the task.

Numeracy Screener. SK students completed the Numeracy Screener tasks. The Numeracy Screener (see Figure 2) is a comprehensive paper-pencil assessment tool consisting of different tasks to measure students' numerical magnitude processing and spatial reasoning (Nosworthy et al., 2013). These tasks were developed based on evidence from developmental psychology research showing strong relationships between these cognitive skills and arithmetic achievement (Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Nosworthy et al., 2013), but the tasks themselves are not necessarily aligned to current kindergarten curriculum expectations. The Screener booklet, including instructions and stimuli, is available on the Open Science Framework. In this cohort, four tasks were included.

1. Number Comparison Task (NCT; 72 items): Students identify which of two digits is bigger
2. Number Ordering Task (NOT; 48 items): Students determine whether three numbers are in the correct order
3. Shape Composition Task (SCT; 20 items): Students identify the shape that would result from putting two 'pieces' together. This task is an adapted version of the widely used Children's Mental Transformation Task (CMTT) developed by Levine, Huttenlocher, Taylor, and Langrock (1999)
4. Mental Rotation Task (MRT; 16 items): Students identify which two images of animals are "perfect matches" if rotated. This task was designed by Neuburger, Jansen, Heil, and Quaiser-Pohl (2011)

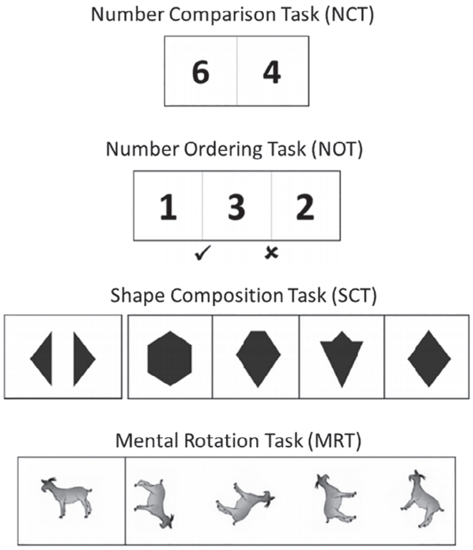

Fig. 2. Sample items from the Numeracy Screener.

Each task is timed, and students have 2 min to respond to as many questions as possible (practice trials including educator instruction and feedback about students' responses are completed first). The dependent variable is the number of items correct for each sub-test. Percentile ranks were calculated so that each student's scores could be considered relative to their peers and enable the school staff to identify students with lower numeracy abilities to target with additional instruction.

Self-Report Data

Self-report data were collected via online surveys to address the research questions. These included a *Post-implementation Survey* for educators to reflect on the process of implementing the assessment tools and *Final Feedback Survey* on educators' experiences with using the results of the assessments in their classroom (see Supporting Information). Most survey items were measured with a Likert scale and analyzed quantitatively. We also collected some open-ended feedback and grouped the responses into common themes that we identified in the data. We then quantified the frequencies of responses for each theme.

Procedure

Educators integrated the Give-N task into the regular classroom environment/curriculum for the JK students and had release time to administer the Numeracy Screener with SK students in a quiet space. The main purpose of the assessments was to enhance student achievement through collaborative inquiry, and thus were conducted in accordance with the Ontario Education Act.

Educators were asked to return the assessment booklets to Curriculum Services within a few weeks. Data were then entered, scored, and analyzed by a research assistant. The curriculum consultant confirmed with educators the names of their JK students who were not CP-knowers and shared the names of their SK students who were in the bottom 10th percentile compared to their peers. This informed the educators of which students needed support in each area addressed in the assessment tools (i.e., counting, number sense, and spatial sense). At this point, the curriculum consultant shared a booklet with mathematical games and activities that the educators were encouraged to implement in their classrooms and track via learning logs (See Supporting Information).

Soon after the assessments, all educators were invited to provide feedback on the implementation process through the Post-implementation Survey. Staff were also encouraged to fill out the final feedback survey in June.

Results

Student Scores on the Assessments

Based on the results of the Give-N task, 78 JK students, roughly 20%, were identified as not yet having reached CP-knower level. SK students' scores on the Numeracy Screener were analyzed to identify students who scored in the bottom tenth percentile on each subtest. Data from the Mental Rotation Task were not interpretable due to floor effects.

Survey Data from Educators

For both assessment tools, most participating educators indicated agreement with most statements regarding readiness for implementation, experiences with administering the tasks, and perceived usefulness of the results. Figure 3 displays the percentage of respondents who strongly agreed or agreed with statements about the Give-N task and Numeracy Screener. Despite some challenges and suggestions for improvements mentioned by participants, respondents indicated that they felt comfortable administering both the Give-N and Numeracy Screener tasks with students, and that implementation went smoothly. All respondents indicated that they felt that students appeared engaged with the Give-N task, whereas only 69% indicated that students were engaged with the Numeracy Screener tasks. Thematic analysis of the open-ended feedback about implementing the Numeracy Screener suggested that the tool's format (i.e., paper-pencil) and instructions presented the biggest challenges for the students, as summarized in Table 1. Additionally, several respondents indicated that the Numeracy Screener took longer to administer than expected. Some educators suggested that it took extra time to explain the

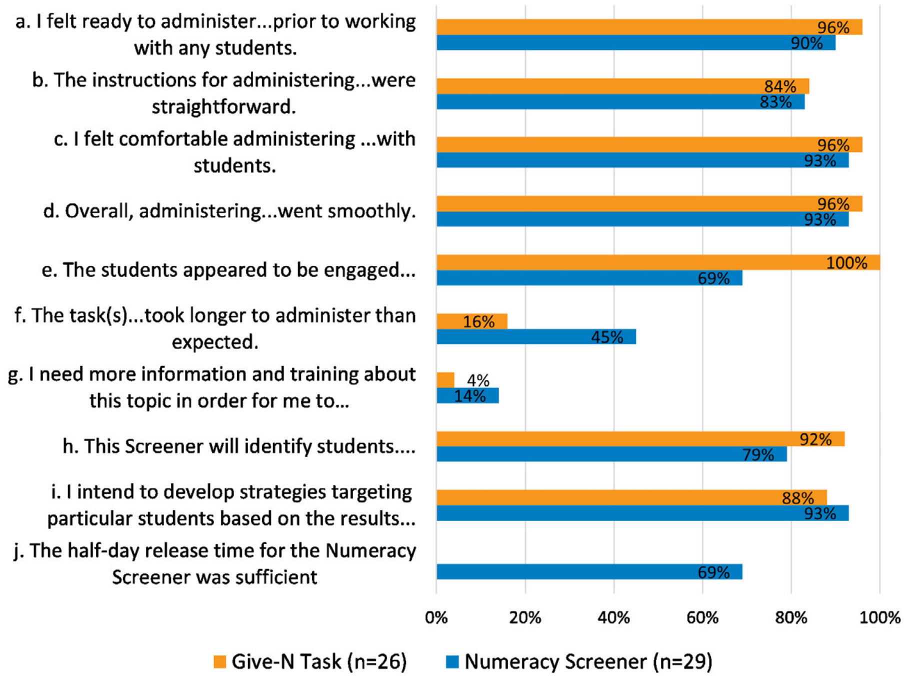

Fig. 3. Agreement (% of 'Strongly Agree' or 'Agree' ratings) about statements regarding implementing the Give-N Task and the Numeracy Screener (Post-Implementation Survey—Cohort 1).

instructions and do the practice trials for each task, and others implied feeling rushed to complete everything during the time provided.

At the end of the school year, educators ($N = 17$) provided feedback on how useful they found the tools in identifying specific mathematical needs of students at risk of falling behind. As shown in Figure 4, the highest proportion of respondents rated both tools as "very" useful (53% for Give-N and 47% for the Numeracy Screener). Open-ended explanations from around three quarters of respondents ($n = 12$) also suggested that the tools helped to identify the students in need of support and areas in which they need support (e.g., *"Clearly identified students who required intervention in number recognition, counting and one to one correspondence"*). Additionally, 38% of respondents mentioned that the tool(s) helped with determining changes to their teaching practice (e.g., "*They helped me see who needed more work on Numeracy. It helped to shape the types of 'math games' we put out.*"). Feedback from a smaller proportion of respondents (31%) also indicated that the students struggled with the paper-pencil or timed format, or the tasks themselves (e.g., "*… thought that the formal 'test' was not a great tool, as students have never completed a timed test before …*" and "*Some of the tasks were extremely difficult for the students (rotation and number ordering)*"). In general, however, the tools appeared to achieve their purpose of identifying specific students/needs on which to focus.

Furthermore, thematic analysis of the question asking educators how they used the assessment results in their classroom showed that most respondents (79%) mentioned that they informed instruction/activities for the students (e.g., "*I used the results to set out provocations and activities for these students to help them build their skills and hit targeted goals throughout the term*"). Around half the respondents mentioned using the assessment results to group students and inform instruction/activities (e.g., "*We made groups based on the results and worked on spatial reasoning and number order as needed*").

Discussion

Year 1 of this inquiry revealed that kindergarten educators found the assessment tools reasonably easy to implement and informative for their practice. These findings suggest that the assessment results can help to determine changes needed in teaching practices to better support particular students become more competent in specific mathematical skills. Educator feedback also highlighted aspects of the implementation of the tools that could be improved. For example, the educators felt that students struggled with the paper-pencil format of the Numeracy Screener, and both the self-report and student assessment data suggested limited utility of one of the spatial tasks (MRT). Educators also observed that this task was difficult for some students and perhaps students did not understand that they had to select

Table 1
Main Themes from Cohort 1 Post-Implementation Survey Responses on Implementing the Give-N Task and Numeracy Screener

	Give-N Task	*Numeracy Screener*
What worked well	Engaging materials/format (e.g., "*The frog had a silly voice and talked to the students, so they thought it was funny and were engaged. They also felt like they were playing as opposed to being tested/learning*") (16 responses) Easy to administer (e.g., "*Simple. Easy to administer. Yielded very good results for planning next steps*") (7)	Implementation Setting (quiet, one-on-one) (e.g., "*It was easily implemented due to the fact that we were in the library with the children one on one not worrying about the classroom")* (11 responses) Flexibility in Instructions/Format (e.g., using highlighters) (9) Easy to administer (7) Implementation supports/release time (5)
Challenges	Complications with materials/instructions (e.g., "*I found that the students wanted to put the cubes in their hand to give to the puppet, instead of push them toward the puppet. This became a problem with the higher numbers and then they would drop the cubes and lose count*.") (8) Implementation setting (distracting) (e.g., "*Needed quiet space for concentration*") (5)	Format/instructions challenging for students * *(*e.g., "*Students had difficulty properly marking their papers with their answers. It took extra time for them to think about how to mark their answer and they appeared confused at times.*") (16) Time (e.g., more needed; took longer than expected) *(*e.g., "*We could not finish the screener tasks in the time that was allotted*") (9)
Suggestions for Improvements	Solutions re: giving the puppet cubes without dropping them (e.g., "*Add to instruction, feed the shark one piece at the time, or put the pieces in a group then feed the shark*.*")* (5)	Change format/instructions (e.g., "*Maybe somehow make it like they Year Ones, more game based and not as rigid, they are not used to that or worksheets.*"; * "*Electronic iPad app would allow the educator to score directly in a digital format where the results can be tabulated more efficiently*.") (8) Time and timing (e.g., "*More time to be released*"; "*starting in September or early October*.") (7)

two responses, unlike in the other tasks. These results highlighted important recommendations for the project team to consider in the next iteration(s) of the project; these recommendations are described below.

COHORT 2

Method

Participants

Fourteen kindergarten classes and 28 educators participated in year 2 of the project. Again, teachers and ECEs received in-service training on administering and scoring the Give-N task and the Numeracy Screener, though task implementation may have been divided among the educators in each classroom. Altogether, 189 JK students and 165 SK students completed the relevant screening tool ($N = 354$). Self-report data were collected from up to 19 educators (68% response rate).

Materials

Give-N Task. The task was the same as in Cohort 1. Changes were made to improve the training for educators in administration and scoring. A member of the research team also checked educators' scoring sheets for accuracy.

Numeracy Screener. Only two subtests of the paper-pencil version of the Numeracy Screener were used with students: the NCT and NOT. Items from the SCT were administered electronically using Qualtrics software on iPads.

Self-Report Measures. Educators completed Post-implementation and Final Feedback Surveys, as with Cohort 1.

Procedure

The procedure in year 2 was like that of year 1, except it incorporated several implementation-related recommendations gathered from Cohort 1. For instance, assessments were completed by the end of October instead of November/December. Additionally, although educators continued to use paper-pencil versions of the two number-sense tasks, they were given extra training and release time to do their own immediate scoring rather than wait for a research assistant to share results. Using data-driven benchmarks from the first cohort, educators could immediately identify if each student was likely to require more targeted supports or activities introduced in the in-service. As performed in the first year, educators provided feedback in the weeks after implementation ended and again at the end of the school year.

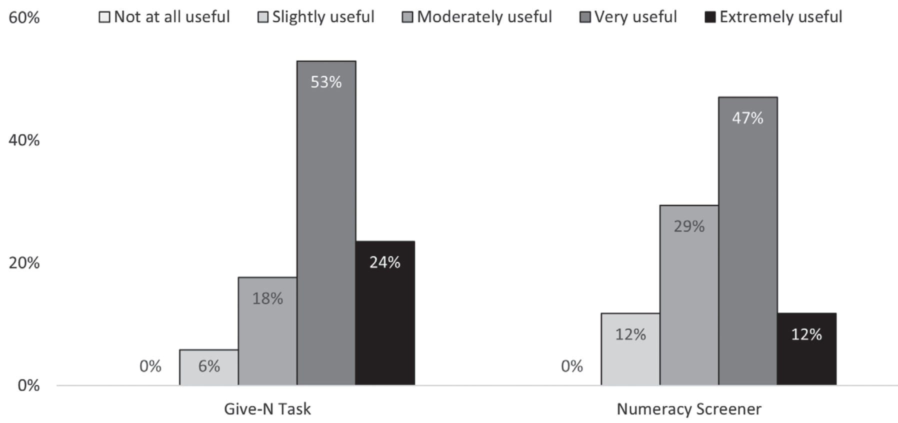

Fig. 4. Proportion of the educators' ratings of usefulness of the Give-N Task and Numeracy Screener (Final Feedback Survey—Cohort 1).

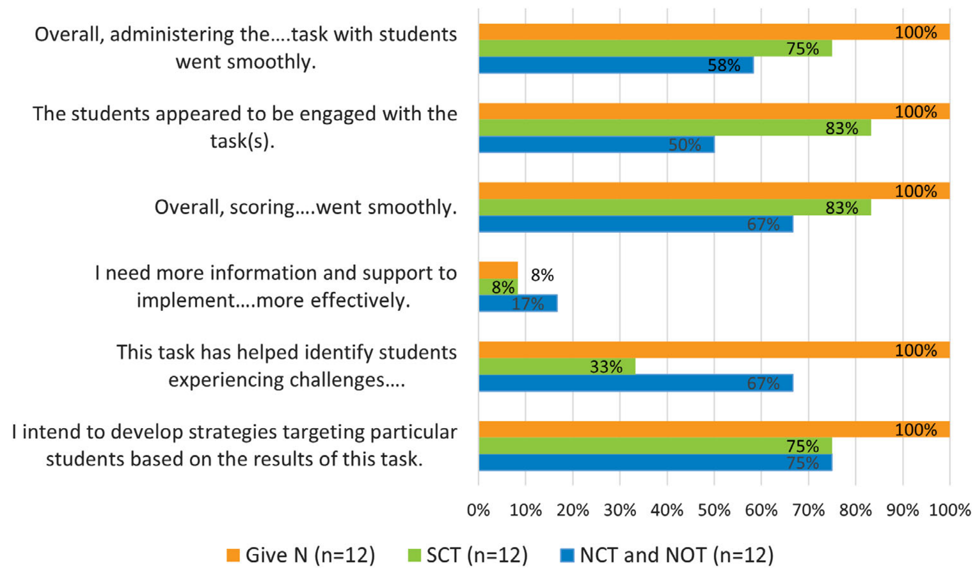

Fig. 5. Agreement (% of 'Strongly Agree' or 'Agree' ratings) about statements regarding implementing the Give-N Task and the Numeracy Screener's spatial (SCT) and number (NCT and NOT) tasks (Post-Implementation Survey—Cohort 2).

Results

Student Scores on the Assessments

Through the Give-N task, around 71 JK students (38%), were identified as not yet having reached CP-knower level. SK students' scores on the Numeracy Screener tasks identified students in the bottom tenth percentile, according to standards calculated based on the data from the previous cohort.

Survey Data from Educators

On the Post-Implementation Survey, all educators who responded about the Give-N task ($N = 12$) agreed with all statements regarding positive experiences with the implementation. Specifically, there was 100% agreement that administering the task went smoothly, students appeared to be engaged, and the scoring went smoothly overall.

While most respondents also indicated agreement with the items about the Numeracy Screener tasks (SCT and NCT/NOT), fewer educators indicated positive feedback compared with the Give-N task. Overall, ratings were lowest for the paper-based NCT/NOT—especially regarding perceptions of student engagement and ease of implementation (see Figure 5). Additionally, when asked about the format of the Numeracy Screener tasks, two-thirds of respondents indicated that the tablet format was easier to implement than paper-pencil. This was also reflected in responses to the open-ended question asking for suggestions to improve the NCT/NOT implementation process: (e.g., "*The children are not familiar with paper and pencil tasks. Less pages to implement and less items per page …*"). Three educators also mentioned some difficulties with the NOT (e.g., "*The number sequencing was extremely difficult for the majority of*

Table 2
Summary of Cohort 2 Post-Implementation Survey Responses on Suggestions for Improving the Process of Implementing the Give-N Task and Numeracy Screener Tasks

Question	*Give-N task*	*Numeracy Screener: Shape Composition Task*	*Numeracy Screener*
Do you have any specific suggestions to Improve the process of implementing the [Task]? If so, please enter them here:	None, positive feedback (e.g., "*The kids loved it and it was a quick way to see which students would benefit from some more activities like this to help build there (sic) number sense*" (2 responses)	Challenging—make task less difficult or resolve confusing layout/format on screen (e.g., "*Since the children are not familiar with this task, more time would be needed.*"; "*The students would get confused as to which question they were attempting to answer as when scrolling you can see multiple. It slowed down the process.*") (6 responses)	Change paper-pencil format (e.g., "*The children are not familiar with paper and pencil tasks. Less pages to implement and less items per page. Some children had a difficult time focusing*.") (4 responses) Number Ordering Task difficult, Number Comparison Task favored more (e.g., "*The Number Ordering Task was the MOST difficult of all of the screeners to take, administer and score. The students had a hard time comprehending what was being asked of them and got caught up in the Check Mark and X....The Number Comparison was much easier to manage both for educators and students, although drawing a line through the number was confusing for them*." (3 responses)

them. They didn't understand how to 'order' numbers ... "). Table 2 contains a summary of the open-ended feedback on the Cohort 2 Post-Implementation survey.

As depicted in the bottom two charts of Figure 5, there was a clear indication that educators felt that Give-N task helped identify students experiencing CP challenges. In contrast, only around one-third of educators agreed the SCT helped identify students with spatial challenges, and two-thirds agreed the NCT/NOT helped identify students with number-related challenges. For all three tools, 75–100% of respondents indicated that they intend to develop strategies targeting particular students based on the results of the task.

On the Final Feedback survey, six of the seven respondents (86%) suggested they believed the screeners achieved their intended purpose. For example, the open-ended responses reflected that the screeners helped identify children in need of support (e.g., "*Sometimes too many assumptions are made regarding the capabilities of the students ... This was a good way to determine their abilities.*"; "*Great to identify some students who fly under the radar and can identify a number but don't know order or quantity*"). Additionally, 71% indicated that the screeners helped them determine changes to their teaching practice and thus to group/target students based on their area of need.

Discussion

Overall, the Give-N task and two of the three Numeracy Screener tasks (i.e., SCT and NCT) were perceived as relatively straightforward to implement in the classroom, and educators had the most positive perceptions of implementing the Give-N task. The assessment tasks perceived as most challenging for teachers and students were the paper-based ones, especially the NOT, which resulted in several scoring errors. Post-Implementation feedback also suggested that the modifications made to the format of the SCT (i.e., tablet-based) were beneficial.

These findings suggest that the Numeracy Screener tasks should be completed online, if possible. Especially given the apparent advantages of the tablet-based SCT version (i.e., automatic data entry and scoring; immediate feedback for educators; saved paper/costs; predominantly positive user-feedback) and the noted challenges with the paper-based versions (i.e., perceived student difficulty with paper-pencil tasks NCT and NOT; observed educator error when calculating scores).

Recommendations for the next year were thus focused on further streamlining the tasks administered to students, as well as the in-servicing, implementation, and scoring approaches for the educators. For the Give-N task, for example, suggested improvements included enhancing capacity by having educators watch a video of the task administration in action, and role-play how to respond to the students and record the scores, using the updated implementation scripts/guides and response sheets. The school board acted on these recommendations by updating the capacity-building sessions/resources and by creating their own digital version of the Numeracy Screener to implement in the third cohort of the project (see Supporting Information).

GENERAL DISCUSSION

Results from 3 years of collaborative inquiry reveal that the Give-N task and Numeracy Screener, assessment tools commonly used in mathematical cognition research, can also be useful in an applied educational setting, particularly early years' classrooms. Participating educators generally found the tools feasible to implement and reported that the assessment results helped them to target math instruction to support students with less mathematical knowledge compared to their peers. Feedback from educators also identified aspects of the assessments that could be improved and highlighted the unique needs of an educational setting. In particular, online administration of the assessment tools using an electronic device such as a tablet is much easier for educators to implement in the classroom and to score. The researchers had not previously adapted the Numeracy Screener to be tablet-based because the paper-pencil version had been used in many countries where there is not always access to tablets or a reliable internet connection. The educators gained valuable insights about how best to use evidence-based assessment tools for their purposes, which are different from the goals of researchers. These findings also provide evidence that collaborative projects between researchers and educators can help address the research-practice gap in education (e.g., Amiel & Tan, 2019).

Limitations

One limitation of this inquiry project is that we were unable to follow students longitudinally to see whether participating in this project was related to final math grades in first and second grade. This is mainly because of privacy considerations for participating students and school board research regulations in Ontario. Another limitation is that survey response rates were low and variable, likely due to the many demands on participating teachers' time. Perhaps giving incentives such as additional release time could help increase response rates in future projects. Feedback from participating educators also highlighted limitations of the Numeracy Screener for kindergarten classroom administration. The Numeracy Screener helped teachers identify number knowledge and spatial skill targets for formative assessment, but there are many other formative assessment tools for early math that have different strengths and limitations. For example, the Number Knowledge Test (Okamoto & Case, 1996) is an in-depth clinical interview style assessment of children's number knowledge that is not speeded. More work is needed to develop and test formative math assessments for kindergarten classrooms.

Directions for Future Research

Future research should explore infrastructure for connecting researchers and educators in cognitive science and education. Getting everyone on board and ensuring sufficient ethical approvals were in place took time and led to delayed testing for the first cohort. Scaling up research-practice partnerships like the one described here requires buy-in and engagement from multiple stakeholders. Researchers should involve educators from the design phase of the project to make sure the project is feasible and aligned with the school board's goals. Future research could investigate ways to deliver the assessments in ways that are easier to integrate into the classroom, particularly for the ordering tasks and spatial reasoning tasks of the Numeracy Screener. Spatial reasoning researchers have recently argued the benefits of assessing spatial thinking skills *within* the learning contexts in which they are used and developed (e.g., during geometry and measurement), urging caution in solely relying on context-free psychological tests, such as mental rotation (Atit, Uttal, & Stieff, 2020). More widespread partnerships could be useful for co-designing and evaluating activities that support cognitive development and early math education.

Implications for Practice

Having the opportunity to implement evidence-based assessment tools with support from researchers is jointly beneficial for educators and school boards, more broadly. It is helpful to also have a designated internal researcher or data champion act as an intermediary between the educators and scientists. In this case, the school board's research team was instrumental in facilitating the collaborative inquiry project with the school board and the research lab. It is also useful to have ongoing support from Curriculum staff on introducing intentional mathematical learning opportunities to students. Moreover, members of the research lab provided workshops for educators, which was an added benefit of the partnership (Christianakis, 2010). Through this professional development and collaboration, participating educators learned about the importance of number knowledge and spatial skills for early math learning. In turn, the participating researchers gained practical considerations for improving their own research practice. Future collaborative inquiry might explore effects of the early assessments and follow-up strategies on student achievement.

CONCLUSION

The implementation of the early numeracy assessment tools in the school setting and accompanying evaluation would not have been possible without effort and engagement from both members of the school board and the research lab. Participants with multiple perspectives gained shared insights into early math education and cognitive development and obtained tangible direction for future research and implementation. It is clear from this work that evaluating the implementation and usage of laboratory-developed mathematical cognition tools has the potential to enhance the implementation process in future, as well as provide important practical feedback to the researchers as to how their assessment tools work "in the wild". The success of this partnership highlights the potential benefits of transdisciplinary collaboration for addressing the gap between research and practice in education.

Acknowledgments—We would like to thank all of the educators and students who participated in this research. This work was supported by a Brain Canada and NeuroDevNet Developmental Neuroscience Training award to R.M., as well as operating grants from the Natural Sciences and Engineering Council of Canada (NSERC), the Canadian Institutes of Health Research (CIHR), the Canada Research Chairs Program, an E.W.R. Steacie Memorial Fellowship from NSERC to D.A.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1 Supporting Information

NOTE

1 Designated ECEs have a diploma in Early Childhood Education and are members of the College of Early Childhood Educators but are not qualified as a teacher. They are part of a two-educator team consisting of one ECE and one certified teacher who co-deliver the full-day Kindergarten program.

REFERENCES

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. *Educational Psychology Review*, *24*(3), 437–448.

Amiel, J. J., & Tan, Y. S. M. (2019). Using collaborative action research to resolve practical and philosophical challenges in educational neuroscience. *Trends in Neuroscience and Education*, *16*, 100116.

Ansari, D., & Coch, D. (2006). Bridges over troubled waters: Education and cognitive neuroscience. *Trends in Cognitive Sciences*, *10*(4), 146–151. http://doi.org/10.1016/j.tics.2006.02.007

Atit, K., Uttal, D. H., & Stieff, M. (2020). Situating space: Using a discipline-focused lens to examine spatial thinking skills. *Cognitive Research: Principles and Implications*, *5*, 19. https://doi.org/10.1186/s41235-020-00210-z

Bruce, C. D., Davis, B., Sinclair, N., McGarvey, L., Hallowell, D., Drefs, M., … Woolcott, G. (2017). Understanding gaps in research networks: Using "spatial reasoning" as a window into the importance of networked educational research. *Educational Studies in Mathematics*, *95*(2), 143–161. http://doi.org/10.1007/s10649-016-9743-2

Campbell, T., & Parr, M. (2015). *Mobilizing research into practice in meaningful ways*. Literacy and Numeracy Secretariat. Retrieved from http://www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/WW_MobilizePractice.pdf

Christianakis, M. (2010). Collaborative research and teacher education. *Issues in Teacher Education*, *19*(2), 109–125. Retrieved from https://files.eric.ed.gov/fulltext/EJ902678.pdf

Donohoo, J. (2013) *Collaborative inquiry for educators: A Facilitator's guide to school improvement*. Thousand Oaks, CA: Corwin.

Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In G. J. Duncan, & R. J. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances*. (pp. 47–69). New York, NY: Russell Sage.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis) alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, *35*(2), 157–178. https://doi.org/10.3102/0162373712461850

Ginsburg, H. P. (2009). The challenge of formative assessment in mathematics education: Children's minds, teachers' minds. *Human Development*, *52*(2), 109–128.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. Social policy report. *Society for Research in Child Development*, *22*(1), 1–24.

Hawes, Z., Nosworthy, N., Archibald, L., & Ansari, D. (2019). Kindergarten children's symbolic number comparison skills relates to 1st grade mathematics achievement: Evidence from a two-minute paper-and-pencil test. *Learning and Instruction*, *59*, 21–33.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, *35*(4), 940–949.

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science*, *17*(5), 714–726.

Maloney, E. A., Converse, B. A., Gibbs, C. R., Levine, S. C., & Beilock, S. L. (2015). Jump-starting early childhood education at home: Early learning, parent motivation, and public policy. *Perspectives on Psychological Science*, *10*(6), 727–732. https://doi.org/10.1177/1745691615607064

Massonnié, J., Frasseto, P., Mareschal, D., & Kirkham, N. Z. (2020). Scientific collaboration with educators: Practical insights from an in-class noise-reduction intervention. *Mind, Brain, and Education*, *14*, 303–316. https://doi.org/10.1111/mbe.12240

Merkley, R., & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: Evidence from brain and behavior. *Current Opinion in Behavioral Sciences*, *10*, 14–20. http://doi.org/10.1016/j.cobeha.2016.04.006

Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2011). Gender differences in pre-adolescents' mental-rotation performance: Do they depend on grade and stimulus type? *Personality and Individual Differences*, *50*(8), 1238–1242.

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS One*, *8*(7), e67918.

Okamoto, Y., & Case, R. (1996). II. Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development*, *61*(1–2), 27–58.

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology*, *105*(2), 453.

Sarnecka, B. W., Negen, J., & Goldman, M. C. (2018). Early number knowledge in dual-language learners from low-SES households. In *Language and culture in mathematical cognition*. (pp. 197–228). Cambridge, MA: Academic Press.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Youmans, S., Coombs, A., & Colgan, L. (2018). Early childhood educators' and teachers' early mathematics education knowledge, beliefs, and pedagogy. *Canadian Journal of Education*, *41*(4), 1080–1104.